Research Article | Vol. 14, No. 6 / 1 Jun 2023 / Biomedical Optics Express 2756
Biomedical Optics EXPRESS

Visual inertial odometry enabled 3D ultrasound and photoacoustic imaging

Deeksha M. Sankepalle,[1] Brian Anthony,[2] and Srivalleesha Mallidi[1,3,*]

[1] *Department of Biomedical Engineering, Tufts University, Medford, MA, 02155, USA*
[2] *Institute of Medical Engineering and Sciences, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA*
[3] *Wellman Center for Photomedicine, Harvard Medical School, Boston, MA, 02115, USA*
[*] *Srivalleesha.mallidi@tufts.edu*

Abstract: There is an increasing need for 3D ultrasound and photoacoustic (USPA) imaging technology for real-time monitoring of dynamic changes in vasculature or molecular markers in various malignancies. Current 3D USPA systems utilize expensive 3D transducer arrays, mechanical arms or limited-range linear stages to reconstruct the 3D volume of the object being imaged. In this study, we developed, characterized, and demonstrated an economical, portable, and clinically translatable handheld device for 3D USPA imaging. An off-the-shelf, low-cost visual odometry system (the Intel RealSense T265 camera equipped with simultaneous localization and mapping technology) to track free hand movements during imaging was attached to the USPA transducer. Specifically, we integrated the T265 camera into a commercially available USPA imaging probe to acquire 3D images and compared it to the reconstructed 3D volume acquired using a linear stage (ground truth). We were able to reliably detect 500 μm step sizes with 90.46% accuracy. Various users evaluated the potential of handheld scanning, and the volume calculated from the motion-compensated image was not significantly different from the ground truth. Overall, our results, for the first time, established the use of an off-the-shelf and low-cost visual odometry system for freehand 3D USPA imaging that can be seamlessly integrated into several photoacoustic imaging systems for various clinical applications.

1. Introduction

Photoacoustic imaging (PAI) is a rapidly developing non-invasive imaging modality whose contrast depends on the tissue optical absorption properties. PAI has been employed in a wide range of applications from cancer [1–3] to cardiovascular imaging [4–6]. PAI takes advantage of the photoacoustic effect, in which absorbed photon energy from a pulsed light source produces a rapid thermoelastic expansion and contraction leading to generation of acoustic waves in tissues [1]. The generated photoacoustic signals can be detected by an ultrasound (US) transducer and can be transformed into functional and molecular maps of tissue such as the tumor oxygen saturation [7–13] or biomarker expression [14–19]. Along with the ubiquitously available non-ionizing and non-invasive US imaging, PAI is now poised to join the armory of clinical imaging modalities. As US and PAI share similar receiver electronics, they can also be integrated into a single imaging system termed as "Ultrasound and Photoacoustic (USPA)" imaging as has been demonstrated previously by several groups [8,20–24].

In general, tissue anatomy is gauged by clinicians and sonographers from US B-mode images. Many traditional USPA imaging systems also generate 2D images. The orientation, volume or complex structure of the anatomy is difficult to visualize using just 2D images. The need for reconstructed 3D volumes, along with larger field of view is undebatable as it can aid clinicians to better visualize anatomy and function [25]. Additionally, 3D volumes can help surgeons to ascertain whether a surgical instrument is placed accurately within the region of interest [26].

#489614 https://doi.org/10.1364/BOE.489614
Journal © 2023 Received 10 Mar 2023; revised 4 May 2023; accepted 4 May 2023; published 17 May 2023

Biomedical Optics EXPRESS

Particularly for applications in cancer theranostics and vascular malignancies, 3D USPA imaging has the potential to substantially improve the clinical outcomes [27–29].

There have been several advances recently in 3D US imaging [26,30]. Given the advantages of combining US with the PAI modality, particularly 3D USPA imaging has not been exclusively studied up until recently due to several reasons listed below. First, 2D array transducers can be used to generate 3D images, however they are very expensive, limited for specific organs such as the ring shaped array transducers used for breast imaging [31–33], or the systems are not portable [34]. Second, mechanical translation of the transducer and optical fiber for USPA imaging is accomplished by attaching the integrated probe to a linear stage as been shown in several studies. For example, breast imaging by Nyayapathi et al. [35], preclinical murine tumors imaged by Mallidi et al. with FujiFilm Vevo LAZR-X system [7] or the handheld system proposed by Lee et al. [36] use translational stages to obtain 3D images. Such translation stage-based systems have limited range of motion and are restricted by the range or length of the linear stages being used. Particularly for motion of transducer that is attached to a non-mobile stage, the clinical applications will be limited due to lack of flexibility. Third, fiducial markers such as tattoos have been used for 3D reconstruction of photoacoustic images. For example, Holzwarth et al. suggested an optical pattern be used as a global coordinate system, where a pre-set tattoo-like grids are placed on the region of interest. These high-contrast tattoo grids act as a guide for estimating the position of the transducer in each image [37]. Although this study was able to achieve 3D reconstruction without any modifications to the transducer or imaging equipment, it requires the application of a tattoo grid on the area of interest before imaging, which is not conducive for several clinical applications such as imaging a wound site. Furthermore, there will be a limitation on the area that can be scanned using the technique along with requirement of extensive reconstruction methods for non-linear or curved surfaces. Fourth, 3D imaging was performed with application specific modulation of light delivery, transducer and customized reconstruction using various algorithms; however they are computationally expensive, time-consuming and system specific methodologies [34,38,39]. Lastly, mechanical localizers and robotic arms such as the daVinci robot have been be used for spatially localized USPA imaging [40]. Such systems, though cost-efficient, have limited availability and can be bulky. Overall, there is a need for handheld USPA system, that can be low cost, portable, attachable to any transducer and light delivery system (i.e., be system independent), not limited in range of motion and conducive for both linear and rotational translation (i.e., have six degrees of freedom of movement in 3D space).

In this study we develop and characterize a low-cost 3D USPA imaging system that has all the aforementioned salient features, namely portability, system independence, unlimited scanning range with six degrees of freedom of movement and low cost (<$300). Specifically, a commercially available USPA transducer was coupled with the low-cost, commercially available T265 camera to obtain a portable, freehand 3D USPA imaging probe that can track freehand movements for 3D reconstruction without the use of fiducial markers. The compact size of the T265 camera ($108 \times 24.5 \times 12.5$ mm) and its lightweight nature (55 g) enable us to design an economical clinically translatable handheld 3D USPA imaging probe. The T265 camera consists of an Inertial Measurement Unit (IMU) and two fisheye cameras. A typical IMU unit consists of a tri-axial accelerometer, gyroscope, and sometimes a magnetometer. Algorithms like the Madgwick filter can fuse all three readings to compute a single orientation parameter called a quaternion [41]. Integrating visual data from the fisheye camera using algorithms like Vi-SLAM (visual simultaneous localization and mapping) can further reliably provide information on the true position and linear velocity of the T265 camera. For example, Hausamann et al. used T265 camera to study the natural head motion of a subject while doing simple tasks such as walking, running, jog [42]. In another study, Benjamin et al. utilized a similar sensor to capture the location of the US transducer to estimate the renal volume during a freehand 3D ultrasound scan

Biomedical Optics EXPRESS

of a kidney [43]. Here for the first time, we investigate the utility of the RealSense camera to obtain 3D USPA images where handheld 2D images can be reconstructed into 3D volume from the quaternion information.

2. Methods and materials

2.1. Phantom fabrication

To characterize our imaging system and validate the reconstruction algorithm, several phantoms were utilized. The first phantom was fabricated by fixating a 0.7 mm diameter graphite pencil lead (Pentel, Hi-Polymer super 50HB) in between 3D printed supporting beams inside a box. This box was then filled with water for USPA imaging. The second phantom was made with two hair strands that were ~103 µm in diameter and placed in a 'X' (crisscross) configuration inside a custom 3D printed box filled with water. The phantom was used for characterizing the system for imaging speed, imaging range and resolution. To facilitate handheld imaging, a third phantom was fabricated with two hair samples in a crisscross configuration embedded in gelatin (CAS#9000-70-8, Sigma-Aldrich, St. Louis, Missouri). Briefly, gelatin powder (8% of w/v) was added to boiling water and stirred until the solution was clear. After the gelatin solution reached ~35°C, it was poured into the mold with the hair sample. The final gelatin block had dimensions 11.5 cm x 8 cm x 2.5 cm.

A fourth phantom was fabricated using a SCRIBD 3D stereo advanced drawing pen loaded with Polylactic acid filament (red color) to compare the range of motion of a linear stage to the integrated handheld probe. A blood vessel structure similar to that in a human arm was 3D printed and then embedded in an 8% w/v gelatin mold. Finally, a fifth phantom was fabricated with rat spleen in 8% w/v gelatin mold to quantify volume from the reconstructed 3D USPA images. As the focus of the optic fiber was at 10 mm, the spleen was positioned 10 mm deep in the gelatin phantom.

2.2. Ultrasound and photoacoustic imaging system

Vevo LAZR-X, a multimodality imaging system by VisualSonics (FUJIFILM, Ontario, Canada) with a 21 MHz transducer (MX250S) fitted with optical fiber jacket was used to acquire USPA images. The Vevo LAZR-X system is equipped with a 20 Hz tunable nanosecond pulsed laser. A default illumination wavelength of 750 nm was used for all experiments in this study as the laser had maximal energy output at this wavelength. The fibers focused light at 10 mm from the base of the transducer and hence all regions of interest in the phantoms were positioned to be 10 mm away from the transducer. Unless otherwise mentioned, USPA image acquisition was performed with no persistence, i.e., 5 Hz frame rate. A lightweight 3D printed mount was designed to hold the transducer, optical fibers, and the T265 camera (Figs. 1(A)–1(B)). Together these parts will be referred as the "integrated" probe in the manuscript. The integrated probe also has a handle to enable users to comfortably hold it during the handheld scanning procedure as shown in Fig. 1(A).

2.3. RealSense hardware

The Intel RealSense T265 camera consists of an IMU sensor (3 Degree of freedom, DOF gyroscope 2000∘s range; 200 Hz sampling rate), and 3 DOF accelerometer (± 4 g range; 62.5 Hz sampling rate) and 2 fisheye world cameras (173-degree diagonal field of view, 848 × 800-pixel resolution; 30 Hz sampling rate), which feed into a Vi-SLAM pipeline (Fig. 2). This algorithm fuses accelerometer, gyroscope, and wide-field image data into a 6 DOF estimation of position and orientation of the T265 camera relative to the environment [44]. The data is computed on an onboard dedicated chipset in real-time which is proprietary to Intel Inc.

Biomedical Optics EXPRESS

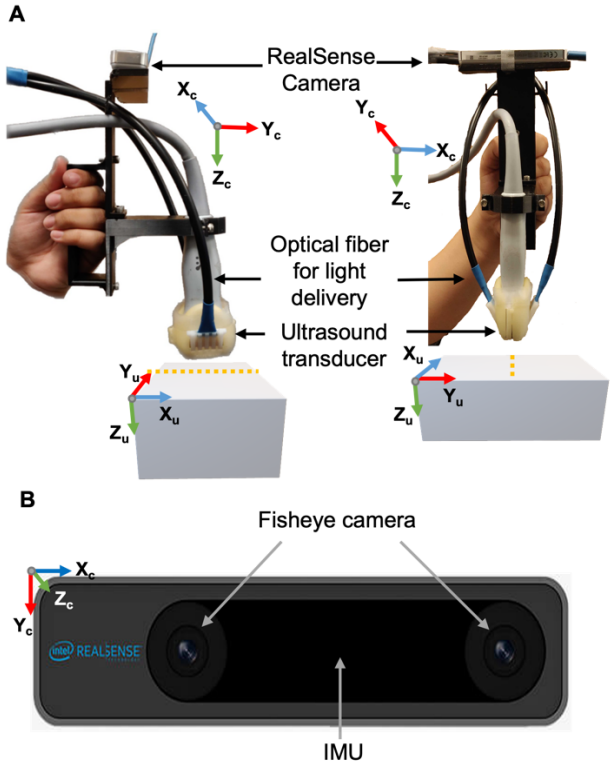


Fig. 1. **A**) Custom 3D-printed handheld probe to housing the US transducer, optical fiber for laser light delivery and Intel RealSense T265 camera. The camera and USPA image co-ordinates are represented as X_c, Y_c and Z_c and X_u, Y_u and Z_u respectively. All axes color coded with blue, red and green for X, Y and Z respectively. **B**) The T265 camera has two fisheye imagers and an integrated IMU where X_c is long axis, Y_c is the short axis, and Z_c is the height of the camera.

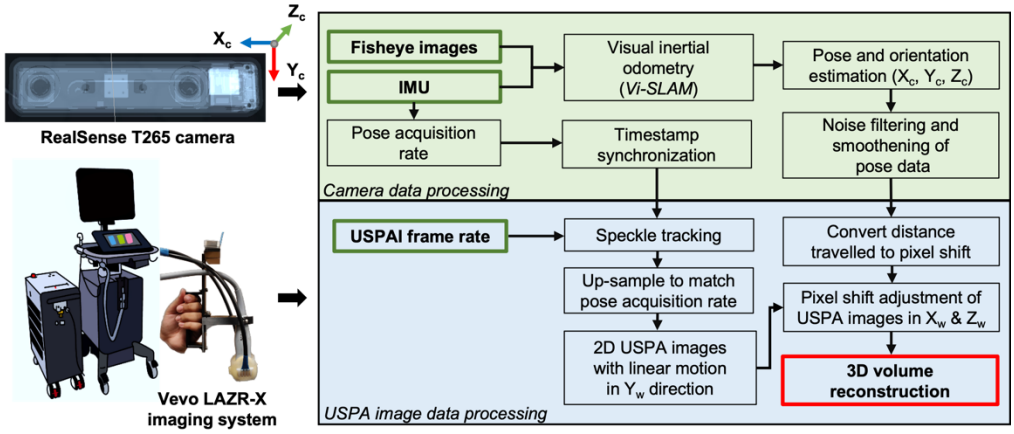


Fig. 2. Flowchart depicting the data processing involved in generating motion-compensated 3D reconstructed USPA images from the T265 camera pose data. Green and red boxes indicate data input and final output stages respectively.

Biomedical Optics EXPRESS

2.4. Image and data processing

The entire data acquisition and image processing flow is represented as a schematic in Fig. 2. The required software packages and wrappers, namely the Intel RealSense SDK (Software Development Kit) and MATLAB wrappers were downloaded from GitHub (GitHub, CA). The Intel RealSense data (.bag files) was recorded on the SDK application provided by Intel. All data and image processing were performed on MATLAB (MathWorks, Natwick, MA). USPA image data from VevoLab was imported into MATLAB for 3D reconstruction. The 3D volumes were then visualized in AMIRA (Thermo Fisher Scientific, Waltham, MA). The pose data from the camera and the USPA images were synchronized based using the timestamps available on the data. The synchronization obtained with timestamps was reconfirmed by identifying motion based on the speckle change in ultrasound images. In static conditions, the US images do not show changes in speckle pattern inside the region of interest in the phantoms. The time of scan and synchronized pose data was obtained from the start and end frames determined by start and end of the speckle change in phantoms.

2.5. Image reconstruction algorithm

Translational pose and camera frame acquisition rate were extracted from the Intel RealSense SDK. The Intel RealSense SDK uses Vi-SLAM to estimate the translational pose and orientation from fisheye images and IMU sensor data. Using ROS (Robotic Operating System) wrappers in MATLAB, the translational pose was imported onto MATLAB. Simultaneously, corresponding original USPA scan data set was imported into MATLAB. The pose data that contained relevant time stamps was then trimmed to match the USPA acquisition time. A smoothing filter (sgolay, degree of polynomial = 0.01) was applied to eliminate jitter noise from the T265 camera. The smoothed pose data was divided by the spatial resolution of the transducer (calculated using Thorlabs NBS 1952) in each axis individually to determine the pixel shift. USPA scans were interpolated to match the frame rate of the T265 camera. Each frame from these scans were then spatially aligned in 3D space, based on the pixel shift previously determined. For visualizing the 3D structures, the voxel sizes were adjusted based on spatial resolution of the transducer.

2.6. Methodology to evaluate minimum detectable distance, accuracy and repeatability by the T256 camera

A linear stage integrated with the FujiFILM Vevo LAZR-X system was used to obtain 3D images that acted as ground truth. To establish the least movement that can be reliably detected by the T265 camera setup, a linear scan of various step-sizes (150, 300, 500, 1000, 1500 and 2000 μm) for a travel range of 3 cm or 10 cm was performed on the hair phantom using the integrated probe. At every step, the persistence (number of frame averages) was set to "Max" to allow the system to average 20 USPA image frames (move the given step size distance, stop, acquire USPA imaging data and continue to move onto the next position. The number of imaging frames acquired were then compared to the number of steps (move-stop-move) detected from the pose data using findpeaks command in MATLAB. This experiment was repeated 3-5 times and the data was used to calculate the accuracy and repeatability of the steps and total distance moved by the camera.

2.7. Methodology to characterize the maximum user speed optimal for 3D reconstruction

To characterize the maximum user speeds optimal for acceptable 3D reconstruction, the transducer was connected to a linear stage (X-LSM, Zaber, Vancouver, Canada) via custom 3D printed holder, to image the hair phantom at varying speeds of 0.5, 1, 2, 3.5, 5 and 10 mm/sec. USPA images were acquired continuously with no persistence. To synchronize the USPA imaging and

Biomedical Optics EXPRESS

the T265 camera data acquisition, recording on the T265 camera was started prior to acquiring USPA images. Furthermore, linear movement on the 3D X-Y-Z linear stage or handheld scan were initiated after a few baseline (no-motion) frames were acquired on the Vevo LAZR-X system.

2.8. User evaluation of the integrated probe

Our next step was to evaluate handheld scans of the rat spleen phantom by six different users with previous experience on USPA imaging, particularly Vevo LAZR-X. During the handheld scan, the users were instructed to move the integrated probe with a constant speed to the best of their abilities. The users were able to watch the USPA images on the Vevo LAZR-X screen while scanning analogous to the clinical imaging scenario. Each user performed scan on the same phantom 3-5 times.

In this study, 3D reconstructed volume of a rat spleen was compared to further establish the performance of our 3D handheld design. A linear translational stage was used to scan the spleen phantom to obtain the 3D USPA images that was used to establish the ground-truth volume. The volume calculated from the handheld imaging by various users, i.e., the 3D images obtained after compensation due to the handheld motion detected by the T265 camera, was compared to the ground truth volume.

3. Results and discussion

3.1. Relationship between the T265 camera and USPA frame axes

We characterized the orientation of the T265 camera with respect to the imaging axes of the Vevo LAZR-X system (schematically represented in Fig. 3(A)). It is critical to gauge the axes transformation between different systems (i.e., between the real world, camera and the imaging frame axes) to enable accurate 3D reconstruction. To avoid any uncertainty in the scan direction required for reconstruction, we chose three main co-ordinates to describe motion in this study. We defined the 'front and back', 'left and right' and 'up and down' motion as X_w, Y_w and Z_w respectively. The world frame (i.e., the real world) acted as the ground reference for the other two co-ordinate systems as shown in Fig. 3(A). The T265 co-ordinate system was defined as X_c (long axis of the camera), Y_c (short axis) and Z_c (height) with camera center as the origin. The orientation of the camera and world frame axes are the same. The 2D USPA image axes are defined as X_u (width of the frame) and Z_u (depth of the frame) with origin at the first pixel. The elevational direction or the axis on which the transducer is scanned for 3D USPA imaging is defined as Y_u axis. In this study, all the axes are color coded where green represented Z axis (up and down in all co-ordinates), red represented Y axis and blue represented the X axis respectively.

To characterize if the T265 camera can record the linear movements of the transducer in the three primary world axes, i.e., along X_w, Y_w and Z_w directions, a simple zig-zag motion was programmed on the linear stage to which the transducer was attached. The schematic representation of the phantom with a 0.7 mm lead (black) in between two supporting beams (grey) was used for USPA imaging as shown in Fig. 3(B)–3(D), where the black arrow depicts the direction of the integrated probe motion and yellow dashed line was the scan length. Figures 3(E)–3(G) exhibit the pose data obtained from the T265 camera when the integrated probe was moved in X_c, Y_c and Z_c direction respectively. Minimal motion was recorded by the camera for axes in which the integrated probe did not move. USPA images of the pencil lead (point source) were continuously acquired during the motion of the integrated probe. The acquired USPA images were displayed as movies in Visualization 1, Visualization 2 and Visualization 3, representing the motion in the X_c (forward and backward), Y_c (left and right) and Z_c (up and down) axes respectively.

We can also notice from the Visualization 1, Visualization 2 and Visualization 3 that the motion recorded by the T265 camera was also observed in the USPA images. Figure 3(H) exhibits 2D

Biomedical Optics EXPRESS

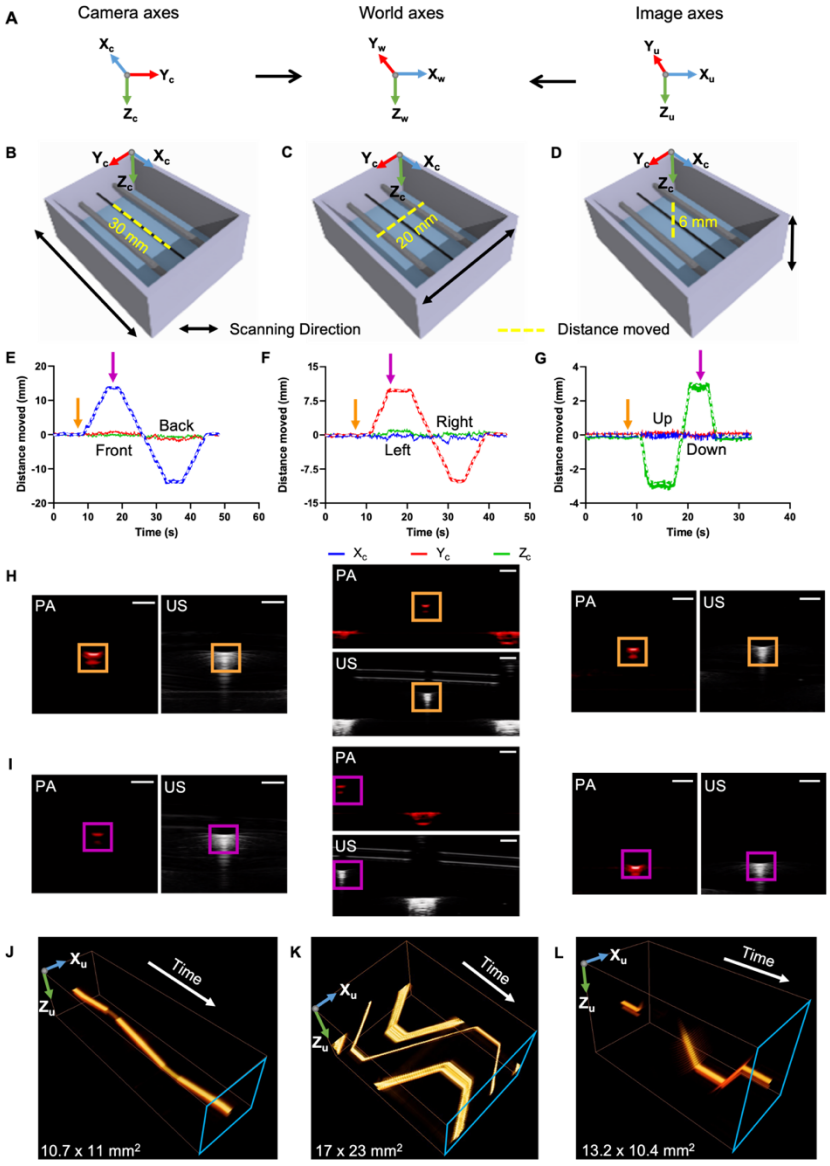


Fig. 3. **A**) Transformation of camera co-ordinates and image co-ordinates to the world co-ordinates. Camera co-ordinates and USPA system co-ordinates are represented as X_c, Y_c and Z_c, and X_u, Y_u and Z_u respectively (All axes color coded with blue, red and green for X, Y and Z respectively). **B-D** Schematic representation of the phantom with a 0.7 mm lead (black) in between two supporting beams (grey) used for USPA imaging. Black arrow depicts the direction of the integrated probe motion and yellow dashed line is the scan length. Enlarged version is provided in Figs. S1-S3. **E-G** represent the translational pose data when the integrated probe is moved in the X_c, Y_c and Z_c direction respectively in the world frame. **H-I** 2D PA and US frames acquired at the positions specified with orange and magenta arrows in **E-G.** The corresponding colored boxes in the 2D images showcase the region of interest i.e., the pencil lead cross-section at locations indicated by the arrows. **J-L** 3D reconstruction of 2D USPA images when camera is moved along the X_c, Y_c and Z_c axes respectively. Red and green axes represent the X_u and Z_u axes of US B-mode image. The white axis represents time. Visualization 1, Visualization 2 and Visualization 3 showcase the images acquired during front-back-front motion, left-right-left motion and up-down-up motion of the integrated probe in the X_c (**J**), Y_c (**K**) and Z_c (**L**) axes respectively.

Biomedical Optics EXPRESS

PA and US frames from the original position (as pointed by the orange arrow in Fig. 3(E)–3(G)) with the lead cross-section highlighted in orange box. Similarly, Fig. 3(I) exhibits 2D PA and US frames at the timepoint specified with magenta arrow in Figs. 3(E)–3(G), with the lead cross-section highlighted in magenta box. For example, in Fig. 3(G), the integrated probe was moved down by 3 mm, moved back up by a total of 6 mm and moved down by 3 mm for it to return to its original position. The USPA images in Figs. 3(H) and 3(l) corroborate with the pose data where the lead cross-section has moved down i.e., the magenta box is lower than the orange box. We clearly see the same motion recorded by the T265 camera in the Z_c axis (Fig. 3(G), (green line)) while no motion was recorded in the X_c (Fig. 3(G), (blue line)) and Y_c axes (Fig. 3(G), (red line)), as expected. Similar motion pattern of "no motion, move certain distance at constant speed, no motion, move back double the distance at constant speed, no motion, and return to original position" was observed for the other two primary axes. Figures 3(J)–3(L) display the 2D USPA images as a 3D image with time as the third axis (represented by the white arrow). Figure 3(J) is the stack of images acquired when probe moved along X_c axis, i.e., along the pencil lead. Hence, we do not see any lateral motion. Figure 3(K) represents the stack of images displayed as 3D image when the probe moved along the Y_c axis. Clearly, we can notice the zig-zag motion along the X_u axis of the images. As noted in Fig. 3(A), motion in the Y_c axis translates to movement in the X_u axis. Figure 3(L) displays the stack of USPA images acquired when the probe was moved up-down in Z_c axis. The fibers attached to the ultrasound probe focus light at 10 mm from the transducer surface, therefore when the pencil lead was too close to the transducer, it was out-of-light focus making the PA signal very weak or absent. However, as the Visualization 3 indicates, the pencil lead can be clearly seen in US images. We were unable to see the pencil lead in photoacoustic image when it is out of laser focus.

3.2. *Evaluating accuracy, repeatability and minimum incremental detectable distance of the T265 camera*

We established the jitter noise of the T265 camera by collecting the pose data when it is not in motion. The unprocessed pose data (non-smoothened) collected over 3 separate days and 3-5 different experiments on each day had approximately a standard deviation of 0.133 mm, 0.199 mm and 0.158 mm in the X_c, Y_c and Z_c directions respectively. As shown Fig. 1(A), the camera was placed facing the ceiling while obtaining the data. If the camera was placed facing a dynamic environment where the participants move around in the room while the camera remained stationary, the standard deviation of the jitter noise was 7.34% and 12.79% higher in the X_c and Y_c directions and 16.09% lower in the Z_c direction. The X_c (front and back) and Y_c (left and right) axes are the predominant scanning directions for USPA imaging and hence we opted for the camera to face the ceiling due to lower jitter in these directions. The movement in the Z_c (up and down) direction is not a predominant scanning direction because it will cause the transducer to move away from the object and create a loss of contact (i.e., acoustic mismatch due to air in between) between the transducer and the object being imaged.

We evaluated the positional accuracy of the camera which is defined as the measure of the error in distance indicated by the camera vs the actual distanced moved, where

$$Accuracy(\%) = 100 - error(\%) \tag{1}$$

$$Error(\%) = 100 * \frac{(distance\ estimated - actual\ distance)}{actual\ distance} \tag{2}$$

For distances ranging 10 mm to 300 mm and camera moving at speeds 1 mm/s -10 mm/s, the average accuracy of the camera for all distances is 98.33%, 95.94% and 97.49% along the X_c, Y_c, and Z_c axes respectively (Table S1). These accuracies are in the similar range as those previously reported with T265 camera in autonomous robotic applications for larger travelling distances [45]. In addition, we notice the accuracy (%) is lower for shorter distances travelled. Given that

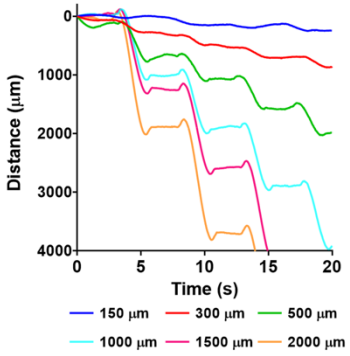
Biomedical Optics EXPRESS

T265 camera was not previously used for photoacoustic imaging applications or smaller travel distances, this is the first report to provide the accuracy for distances in the centimeter range or lower. Specifically in our studies, we noticed accuracy was 96.67% for 10 mm travel distance but accuracy was 98.18% for 300 mm travel distance, along the X_c axis. Fig. S5 and Table S1 clearly shows the accuracy (%) is higher with increased travel distance in all the three axes directions. Furthermore, motion along the Y_c axis (short axis of the camera) had the lowest accuracy as was also previously observed in large scale applications [45].

In addition to the accuracy, we also calculated the positional repeatability of the camera, where repeatability is defined as the extent to which successive attempts to move to a specific location vary in position i.e., the error in the pose data in reporting the position time after time. The position reported by the camera pose data was compared to the position set on the linear stage to calculate the error. Standard deviation of the error is reported as the repeatability of the camera. It has to be noted that the linear stage inherently has an accuracy of 20 µm and repeatability error of ~3 µm. Ignoring the impact of this error on our results, we noticed over several runs, for several distances, the average repeatability to be 210.1 µm, 79.4 µm and 457.4 µm along X_c, Y_c and Z_c axes respectively (Table S1).

The next study involved evaluation of the minimum distance reliably tracked by the T265 camera. Figure 4(A) shows the smoothened pose data for various step-sizes. The minimum step-size we could reliably differentiate from the background jitter in the pose data was found to be 500 µm as shown in Fig. 4(A), where the steps were clearly identified. Specifically, the accuracy for 500 µm, 1000 µm, 1500 µm, and 2000 µm step sizes was 90.46%, 91.52%, 91.32% and 90.51% respectively and the repeatability was 120.92, 159.84, 131.80 and 135.55 µm respectively. The findpeaks command on MATLAB also identified the accurate number of step-sizes above 500 µm (Fig. 4(B)). For step-sizes less than 500 µm, differentiating and computing number of steps from the pose data was not reliable and did not match the number of USPA frames acquired (Fig. 4(B)). In other words, the accuracy of the camera in detecting these small step sizes of 150 µm and 300 µm was less than 50% and these step-sizes were in the repeatability error range mentioned above.

A



B

Step size (µm)	No of USPA frames (total scan distance)	No of steps from IMU	Accuracy (%)	Repeatability (µm)
150	66 (3 cm)	63	< 50 %	N/A
300	33 (3 cm)	35		
500	61 (10 cm)	61	90.46 %	120.92
1000	30 (10 cm)	30	91.52 %	159.84
1500	20 (10 cm)	20	91.32 %	131.80
2000	15 (10 cm)	15	90.51 %	135.55

Fig. 4. **A)** Various step-sizes recorded on the T265 camera when mounted on to a linear stage with movement in the Y_c axis. The pose data shown represents the smoothened data for each step size after application of Savitsky-Golay filter. **B)** Comparison of step-sizes recorded on the T265 camera with the number of USPA frames acquired during linear move-stop-acquire image-repeat motion in Y_c direction. Fig. S4 represents the raw pose data with an overlay of smoothened pose data.

Biomedical Optics EXPRESS

3.3. Establishing the maximum speed that can be used for linear motion of integrated probe with 20 Hz nanoseconds pulsed laser

In systems such as the FujiFILM Vevo LAZR-X used in this study, the frame acquisition rate is about 5 - 20 Hz. Low frame rates accompanied with high-speed translation motion can lead to low sampling of the object being imaged and therefore an erroneous 3D reconstruction. Imaging systems such as the Vevo LAZR-X system utilize a "move-stop-acquire image-repeat" scanning methodology with the linear translational stage. However, such a scenario with pre-determined step sizes is not possible with free hand imaging. To characterize the maximum speed at which the integrated USPA probe can be used for reliable 3D reconstruction of the object without loss of data, the integrated probe was attached to a linear stage that moved at constant speed. The speed ranged from 0.5 mm/s to 10 mm/s for a fixed travel distance of 30 mm (Fig. 5(A)). A linear correlation analysis was performed on the speeds set on the linear stage and speed calculated from the T265 camera's pose data and $R^2 = 0.986$ was observed (Fig. 5(B)). It can also be noted that speed across other axis was zero (red and green data points in Fig. 5(B)). We also acquired USPA images of hair phantom while translating the integrated probe at various speeds. A higher number of USPA frames at slower speeds were obtained than at higher speeds, as expected. As seen in Fig. 5(C), 3D reconstruction of USPA images at 10 mm/s was missing significant structural information, such as the intersection of the two hair strands. At speeds 5 mm/s or lower, the motion compensated reconstructions were similar to the actual phantom. As USPA images were acquired with 20 Hz frame rate, the minimum distance traversed between two adjacent frames was 250 µm for a 5 mm/s travel speed. Though the distance between adjacent frames is in the range of the elevational resolution of the transducer (~300 µm), it does not satisfy the Nyquist criterion, but can be used for 3D reconstruction with interpolation between frames for qualitative 3D representation. A travel speed of 3 mm/s that generates ~150 µm distance between frames or lower speeds will be required for accurate 3D reconstructions. With availability of pulsed lasers that operate at high pulse repetition frequency, several frames can be acquired satisfying the Nyquist criterion and providing accurate 3D reconstruction of the object being imaged.

3.4. Quantitative comparison of scanning speeds of various users using the integrated probe for reliable 3D reconstruction

Our next step was to evaluate the potential of 3D reconstruction of the gelatin hair phantom (Fig. 6(A)) when imaged by various users that were not pre-trained on holding the integrated probe but were familiar with USPA imaging. The users were instructed to perform multiple scans while looking at the near real time USPA images on the Vevo LAZR-X screen. Scan speeds of all the users calculated from the T265 camera pose data are reported in Fig. 6(B). As can be noted there were inter- and intra-scanning speed differences between users. User 3 has the highest average scan speed whereas the User 2 has the most consistent scan speed. Similar to 3D reconstructions for scan performed by a motor at various speeds (section 3.3(above)), users who scanned at lower speeds (example User 2) were able to capture higher number of USPA image frames while users who moved the integrated probe at higher speeds had low number of USPA frames (example User 3) as expected. Obtaining high number of USPA image frames (low speed while moving the integrated probe) produced a better 3D reconstruction of the phantom than that of 3D reconstruction from users who scanned at higher speed. As shown in Fig. 6(C) and 6(D), when imaged at lower speed by User 2 (~3.5 mm/s) and higher speed by User 3 (14 mm/s) respectively, the 3D reconstruction was better in the former case. Corresponding translational pose data acquired by the T265 camera for the handheld scans shown in Figs. 6(E)–6(F). Clearly, the slope (speed) of the X_c pose data is steeper for the higher speed scan. We also observed that the users were able to maintain constant speed for the duration of the scan for lengths 10-15 cm. If the user cannot maintain a constant scan speed for longer scans lengths, we do not anticipate any limitations for the 3D reconstruction as we use the actual position and orientation data and

Biomedical Optics EXPRESS

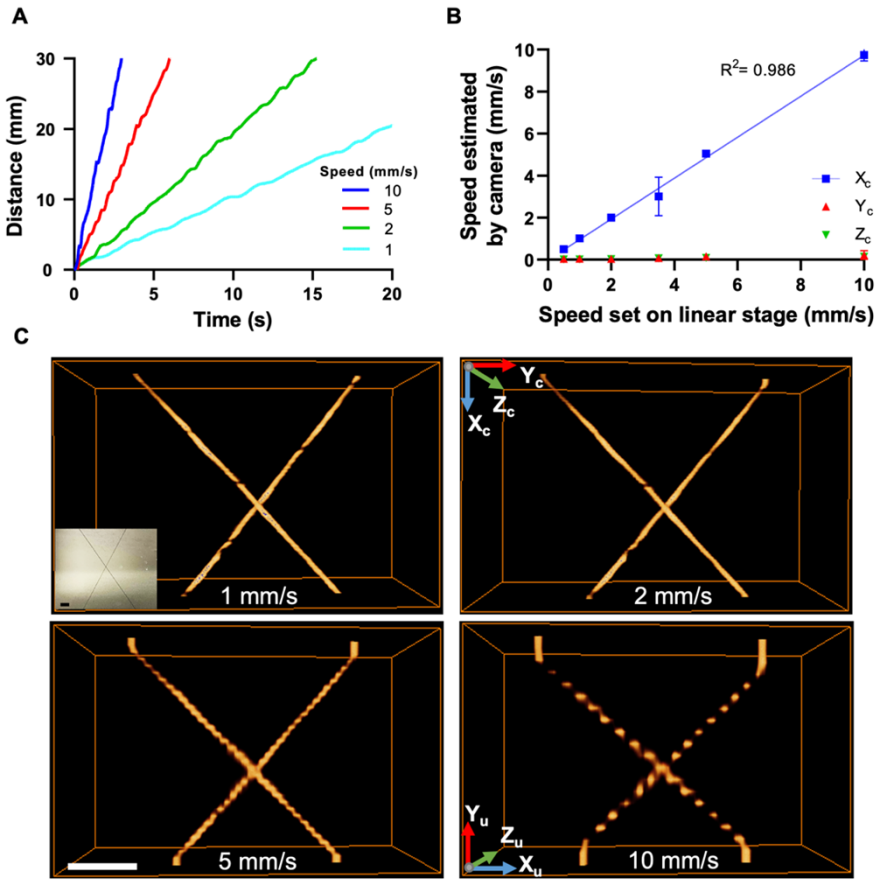


Fig. 5. **A**) Distance travelled by the integrated probe for various speeds programmed on the linear stage. The slope of the pose data provides the speed recorded by the camera. **B**) Comparison of the speed on the linear stage to the speed recorded by the T265 camera (R^2= 0.986). **C**) Motion compensated 3reconstructed PA images for various speeds ranging from 1 to 10 mm/s. Inset: Photograph of the hair phantom (scale bar: 5 mm).

not the speed at a particular time point. Overall for a USPA imaging system operating at 20 Hz frame rate, a scan speed of 5 mm/s or less would be optimal.

3.5. Comparison of the range of motion of linear stage VS handheld imaging probe

The range of motion that can be achieved with linear motor and our handheld probe are compared in Fig. 7. A phantom of approximately 160 mm length was used for this experiment (Fig. 7(A)). Figures 7(B)–7(C) depict the 3D PA and US handheld data, and linear stage. Due to the limited range on the linear stage, only 45 mm scan length was possible. However, with handheld scan, we were able to image the whole length of the phantom, as shown in Fig. 7(B). Clearly, a high visual correlation between the phantom picture and the handheld motion compensated 3D reconstruction can be observed. Although, linear stages with larger range can be purchased, they can be bulky and non-portable. In certain cases, multiple linear stages could be required to capture the whole phantom in 3D, which can significantly increase the imaging time and 3D reconstruction complexity. With our integrated handheld probe, we were able to image the entire phantom in a single scan. Due to this feature, the integrated probe has high potential to image large scan areas making it one of the main advantages of this probe.

Biomedical Optics EXPRESS

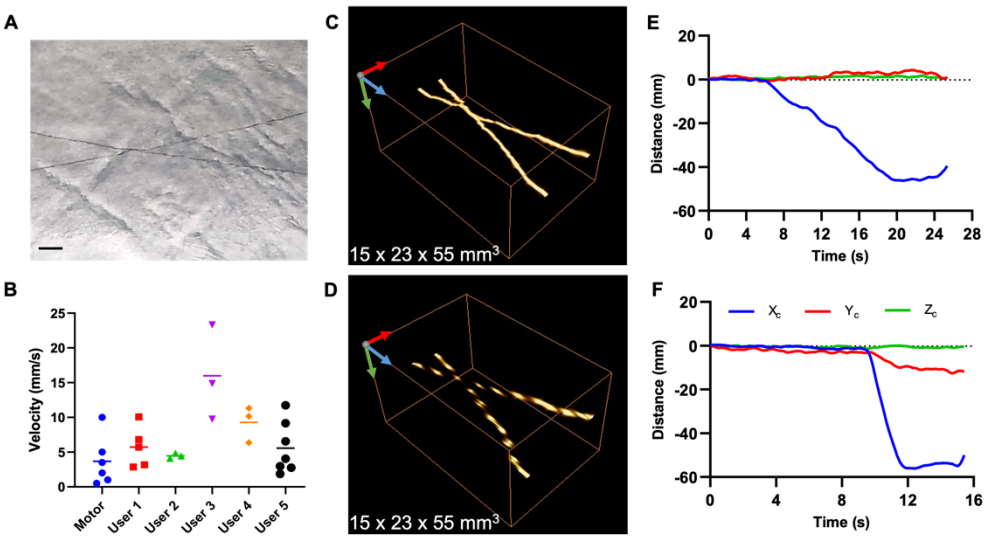


Fig. 6. **A**) Tissue mimicking phantom with hair placed as 'X' in gelatin was used for freehand USPA scan (scale bar: 5 mm). **B**) The graphs represent the speed measured by the T265 camera for users performing a freehand USPA scans. Average speed of each user is represented by the horizontal line. **C-D** 3D reconstruction of motion compensated USPA scans along X_W axis (blue arrow) when imaged by a user at lower speed (User 2: ~3.5 mm/s) and higher speed (User 3: 14 mm/s) respectively. **E-F** Corresponding translational pose data acquired by the T265 camera for the handheld scans shown in **C** and **D**.

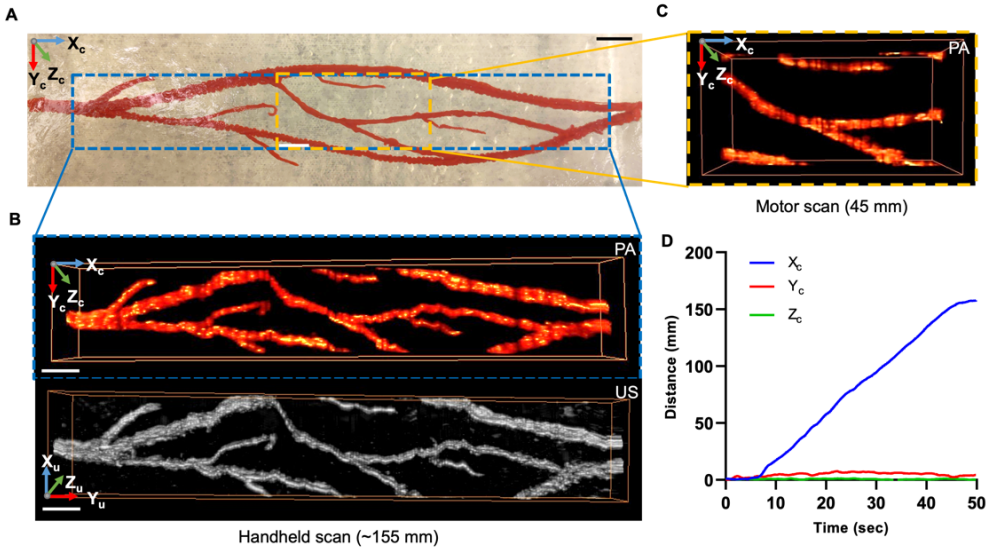


Fig. 7. **A**) Photograph of 3D printed blood vessel phantom. **B**) 3D reconstructed PA and US images of the phantom (imaged area highlighted in blue rectangle; ~155 mm) using our handheld integrated probe. PA and US images were acquired simultaneously using Vevo LAZR-X system. Scale bar: 10 mm. **C)** 3D reconstruction of blood vessel phantom (as shown in A) using linear motor. The maximum scan length achievable on this motor was 45 mm (highlighted in yellow rectangle). **D)** The translational pose data of the corresponding handheld scan**.**

Biomedical Optics EXPRESS

3.6. Volume estimation from the images acquired with our integrated T265 and USPA probe

Post characterization of the 3D handheld integrated USPA imaging probe, we explored its ability to estimate the true volume of a tissue with the pose data acquired from the T265 camera. The handheld reconstructed volume was compared to the volume calculated from linear stage scan. The stack of 2D images during the handheld scans are referred as 'motion uncompensated' data, whereas the 3D reconstruction of these 2D interpolated images using the pose data is referred as 'motion compensated' data. A rat spleen *ex vivo* embedded in a gelatin phantom (Fig. 8(A)) was imaged using the integrated probe. Figures 8(B)–8(C) summarize the volume analysis on reconstruction of freehand scans (6 different users) on the rat spleen phantom. The ultrasound and photoacoustic 3D reconstructions of the spleen in three different views (top, side and front) was displayed in Fig. 8(D). We can clearly notice that the top view of the spleen shown in Fig. 8(D) matches with the ground truth and motion compensated 3D images but not the uncompensated handheld scanned image. We can also notice that the uncompensated handheld images (Fig. 8(D) middle panel) are compressed in the Y_u axis due to the unavailability of the scan length. This suggests that uncompensated 3D reconstruction underestimated the volume of the spleen. Representative pose data from a handheld scan of the spleen phantom is shown in Fig. 8(B). We can clearly observe that there was no motion up until 9 seconds after the start of the acquisition. After 9 seconds, there is translation in the X_c axis (Y_u in the USPA image frame). Upon motion compensation using the pose data from the T265 camera (Fig. 8(B)), we accomplished a 3D reconstruction of a freehand scan, which is now structurally similar to the ground truth. Volume of the spleen was estimated from the manually segmented USPA images for uncompensated and motion-compensated 3D data using MATLAB segmentation toolkit. As mentioned previously, here we assume the volume estimated from the linear translation stage as our true volume. Percentage difference in volume of the spleen from the ground truth is plotted in Fig. 8(C) for motion compensated and uncompensated images respectively for data obtained by six different users. Clearly the difference in volume between the ground truth and handheld scan is averaging around zero as expected (Fig. 8(C), orange bar). A simple t-test produced p-values < 0.0001, indicating the percentage difference in volume of uncompensated and compensated volumes calculated for all freehand scans is significantly different.

Very recently Jiang et al. have used a GPS based system for 3D photoacoustic imaging using G4 system from Polhemus Inc [46]. The G4 system offers similar features like the T265 camera, i.e., it is portable, scalable and compact (similar size) but the major difference is that the G4 system is a 3-piece electromagnetic tracking system while the T265 camera combines inertial tracking with Vi-SLAM algorithms in one system combinedly referred to as Visual odometry system. Electromagnetic tracking units may experience interference when operating in the vicinity of devices that produce magnetic fields and metal objects present in the rooms can also disrupt the magnetic fields. Jiang et al. have taken additional precautions to avoid presence of magnetic distortion by specifically placing the sensor 8 cm behind the middle line of linear array probe. While the T265 camera is not impacted by electromagnetic distortions, studies have shown that the performance of the T265 camera is impacted by bright light such as sunlight in outdoor environments [47]. However, such bright lights are unusual in a laboratory or a clinical environment, making the visual odometry based freehand USPA imaging a viable technique for three dimensional visualization of tissues as demonstrated by our results.

Biomedical Optics EXPRESS

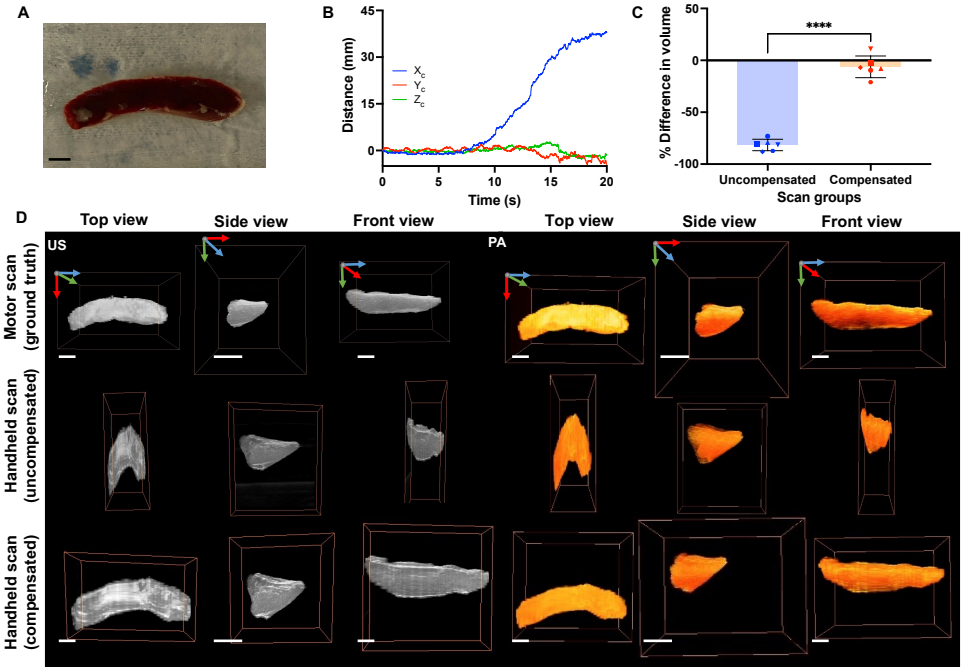


Fig. 8. **A)** Photograph of the rat spleen (*ex vivo*) embedded in 8% gelatin for handheld and motorized USPA imaging. **B**) Translational pose data acquired by the T265 camera for the handheld scan shown in D. **C**) Comparison of volumes estimated from the uncompensated and compensated handheld scans with the ground truth obtained from the linear motor scan. Each symbol in the graph represents a different user conducting the handheld scan (6 users). Percentage difference in volume was computed for all the 3D reconstructed spleen images. A significant difference ($p < 0.0001$) in volume was observed between uncompensated and compensated groups. **D**) 3D US and PA images of *ex vivo* spleen in top (left column), side (center column) and front (right column) view for motorized scan (considered as ground truth) (top row) and handheld uncompensated scan (center row) and compensated scan (bottom row). Scale bar: 5 mm. Coordinates represent X_c (blue), Y_c (red) and Z_c (green).

4. Conclusions and future work

In this study, we developed a low-cost, adaptable, and system independent freehand 3D USPA imaging probe. The handheld probe is a combination of the ultrasound transducer to acquire USPA signals, fiber optics to deliver laser pulses and the Intel T265 camera which consists of two fisheye cameras and an IMU sensor for tracking the probe position. Intel RealSense is primarily used in robotics for localization, where the range of motion is in meters. This is the first time where such cameras are utilized for photoacoustic imaging. While similar IMU based systems were previously used for ultrasound imaging and have been extensively reviewed elsewhere [48,49], here we present the use of visual odometry for the first time for combined ultrasound and photoacoustic imaging. We have evaluated that camera facing the ceiling of the room provides the best option to avoid such distortions due to dynamic environment and real-world clinical rooms and imaging suites have ceilings with railings and other patterns (false ceiling) that can act as fiducial landmarks. If rooms have ceilings that are devoid of patterns or fiducial markers, taping printed patterns on the ceiling could resolve the issue. However, the ergonomics and functionality of the integrated probes in different environments with different ceiling patterns needs further

Biomedical Optics EXPRESS

investigation and are out of the scope of the current work that is focused on demonstrating the feasibility of using visual odometry for combined 3D USPA imaging.

We chose Intel RealSense T265 camera primarily due to its low cost and relatively better performance than other readily available odometers [50]. The T265 camera, being a single unit system, can be attached to any transducer operating at lower frequencies than that used in this study (20 MHz) as the calculated accuracy, jitter noise, repeatability and minimum incremental distance calculated for the current camera are on the order of lateral and elevational resolution of the transducer used in this study. We anticipate that sensors with micrometer range accuracy and precision will be developed in future that can be integrated with such handheld systems while being economical, accurate, portable, and less bulky. The T265 camera provided 6 DOF pose information, i.e., both translational and rotational information of the transducer is provided. In this study we demonstrated the salient features of utilizing T265 tracking camera for linear translation motion and optimized the scanning speed for handheld imaging. Our future studies will involve replicating a true clinical freehand motion that will include both rotation and linear translation. Furthermore, the current commercial USPA imaging system used in this study does not facilitate synchronizing and automating simultaneous acquisition of USPA images and T265 pose data. Our next step is to integrate the T265 camera with USPA image acquisition to provide close to real-time 3D image generation with greater flexibility customized to the clinical or preclinical applications. The major advantage of our integrated probe is that there is no limit to the scan length. A range of few millimeters to several tens of centimeters can be scanned with the integrated handheld probe. With the current frame rate for PA imaging, a scanning speed of 5 mm/s or lower is optimal to get a good 3D reconstruction of the object. With availability of lasers or light sources with high pulse repetition frequency, such as the LED based photoacoustic systems [10,51], we envision the scanning speeds to be higher than those reported in this manuscript. Our future work will involve building integrated systems where real-time feedback will be provided to the user indicating either to slow down or increase the scanning speed to achieve optimal 3D reconstruction of the tissue being imaged. Overall, our results for the first time established the use of an off-the-shelf, low cost visual odometry system for freehand 3D USPA imaging that can be seamlessly integrated to several photoacoustic imaging systems for various preclinical and clinical applications.

Funding. National Institutes of Health funds (S10OD026844 and R01CA231606); Tufts University Clinical and Translational Science Institute funds via NIH UL1TR002544.

Acknowledgments. The authors would also like to acknowledge Dr. Tayyaba Hasan at Massachusetts General Hospital and the members of the integrated Biofunctional Imaging and Therapeutics laboratory (Skye Edwards, Allison Sweeney, Marvin Xavierselvan and Christopher Nguyen) for their useful discussions and support.

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. S. Mallidi, G.P. Luke, and S. Emelianov, "Photoacoustic imaging in cancer detection, diagnosis, and treatment guidance," Trends Biotechnol. **29**(5), 213 (2011).
2. J. Yu, H.N.Y. Nguyen, W. Steenbergen, and K. Kim, "Recent development of technology and application of photoacoustic molecular imaging toward clinical translation," J. Nucl. Med. **59**(8), 1202–1207 (2018).
3. L., Lin and L.V. Wang, "The emerging role of photoacoustic imaging in clinical oncology," Nat. Rev. Clin. Oncol. **19**(6), 365–384 (2022).
4. S. Iskander-Rizk, A.F.W. van der Steen, and G. van Soest, "Photoacoustic imaging for guidance of interventions in cardiovascular medicine," Phys. Med. Biol. **64**(16), 16TR01 (2019).
5. Y. Suzuki, H. Kajita, S. Watanabe, K. Okabe, H. Sakuma, N. Imanishi, S. Aiso, and K. Kishi, "Application of photoacoustic imaging for lymphedema treatment," J. Reconstr. Microsurg. **38**(03), 254–262 (2022).
6. J.L. Su, B. Wang, K.E. Wilson, C.L. Bayer, Y.-S. Chen, S. Kim, K.A. Homan, and S.Y. Emelianov, "Advances in clinical and biomedical applications of photoacoustic imaging," Expert Opin. Med. Diagn. **4**(6), 497–510 (2010).

Biomedical Optics EXPRESS

7. S. Mallidi, K. Watanabe, D. Timerman, D. Schoenfeld, and T. Hasan, "Prediction of tumor recurrence and therapy monitoring using ultrasound-guided photoacoustic imaging," Theranostics **5**(3), 289–301 (2015).
8. S.C. Hester, M. Kuriakose, C.D. Nguyen, and S. Mallidi, "Role of ultrasound and photoacoustic imaging in photodynamic therapy for cancer," Photochem. Photobiol. **96**(2), 260–279 (2020).
9. M. Xavierselvan, M.K.A. Singh, and S. Mallidi, "In vivo tumor vascular imaging with light emitting diode-based photoacoustic imaging system," Sensors **20**(16), 4503 (2020).
10. R. Bulsink, M. Kuniyil Ajith Singh, M. Xavierselvan, S. Mallidi, W. Steenbergen, and K.J. Francis, "Oxygen saturation imaging using LED-based photoacoustic system," Sensors **21**(1), 283 (2021).
11. A. Ron, X.L. Deán-Ben, S. Gottschalk, and D. Razansky, "Volumetric optoacoustic imaging unveils high-resolution patterns of acute and cyclic hypoxia in a murine model of breast cancer," Cancer Res. **79**(18), 4767–4775 (2019).
12. M. Li, Y. Tang, and J. Yao, "Photoacoustic tomography of blood oxygenation: a mini review," Photoacoustics **10**, 65–73 (2018).
13. T.L. Lefebvre, E. Brown, L. Hacker, T. Else, M.-E. Oraiopoulou, M.R. Tomaszewski, R. Jena, and S.E. Bohndiek, "The potential of photoacoustic imaging in radiation oncology," Front. Oncol. **12**, 1 (2022).
14. S. Mallidi, T. Larson, J. Aaron, K. Sokolov, and S. Emelianov, "Molecular specific optoacoustic imaging with plasmonic nanoparticles," Opt. Express **15**(11), 6583 (2007).
15. K.A. Homan, M. Souza, R. Truby, G.P. Luke, C. Green, E. Vreeland, and S. Emelianov, "Silver nanoplate contrast agents for in vivo molecular photoacoustic imaging," ACS Nano **6**(1), 641 (2012).
16. S. Manohar, C. Ungureanu, and T.G. Van Leeuwen, "Gold nanorods as molecular contrast agents in photoacoustic imaging: the promises and the caveats," Contrast Media Mol. Imaging **6**(5), 389–400 (2011).
17. L. Xi, S.R. Grobmyer, G. Zhou, W. Qian, L. Yang, and H. Jiang, "Molecular photoacoustic tomography of breast cancer using receptor targeted magnetic iron oxide nanoparticles as contrast agents," J. Biophotonics **7**(6), 401 (2014).
18. Y. Liu, L. Nie, and X. Chen, "Photoacoustic molecular imaging: from multiscale biomedical applications towards early-stage theranostics," Trends Biotechnol. **34**(5), 420–433 (2016).
19. Y. Zheng, M. Liu, and L. Jiang, "Progress of photoacoustic imaging combined with targeted photoacoustic contrast agents in tumor molecular imaging," Front. Chem. **10**, 1 (2022).
20. A.B. Karpiouk, S.R. Aglyamov, S. Mallidi, J. Shah, W.G. Scott, J.M. Rubin, and S.Y. Emelianov, "Combined ultrasound and photoacoustic imaging to detect and stage deep vein thrombosis: phantom and ex vivo studies," J. Biomed. Opt. **13**(5), 054061 (2008).
21. R. Manwar, K. Kratkiewicz, and K. Avanaki, "Overview of ultrasound detection technologies for photoacoustic imaging," Micromachines **11**(7), 692 (2020).
22. G.S., Sangha and C.J. Goergen, "Label-free photoacoustic and ultrasound imaging for murine atherosclerosis characterization," APL Bioeng. **4**(2), 026102 (2020).
23. Y.H. Wang, A.H. Liao, J.H. Chen, C.R. Wang, and P.C. Li, "Photoacoustic/ultrasound dual-modality contrast agent and its application to thermotherapy," J. Biomed. Opt. **17**(4), 045001 (2012).
24. W. Wei, X. Li, Q. Zhou, K.K. Shung, and Z. Chen, "Integrated ultrasound and photoacoustic probe for co-registered intravascular imaging," J. Biomed. Opt. **16**(10), 106001 (2011).
25. E. Dyer, U. Zeeshan Ijaz, R. Housden, R. Prager, A. Gee, and G. Treece, "A clinical system for three-dimensional extended-field-of-view ultrasound," Br. J. Radiol. **85**(1018), e919–e924 (2012).
26. A., Fenster and D.B. Downey, "3-D ultrasound imaging: a review," IEEE Eng. Med. Biol. Mag. **15**(6), 41–51 (1996).
27. J. Laufer, P. Johnson, E. Zhang, B. Treeby, B. Cox, B. Pedley, and P. Beard, "In vivo preclinical photoacoustic imaging of tumor vasculature development and therapy," J. Biomed. Opt. **17**(5), 056016 (2012).
28. E. Petrova, A. Liopo, V. Nadvoretskiy, and S. Ermilov, "Imaging technique for real-time temperature monitoring during cryotherapy of lesions," J. Biomed. Opt. **21**(11), 116007 (2016).
29. E.V. Petrova, H.P. Brecht, M. Motamedi, A.A. Oraevsky, and S.A. Ermilov, "In vivo optoacoustic temperature imaging for image-guided cryotherapy of prostate cancer," Phys. Med. Biol. **63**(6), 064002 (2018).
30. Q., Huang and Z. Zeng, "A review on real-time 3D ultrasound imaging technology," BioMed Res. Int. **2017**, 6027029 (2017).
31. A. Oraevsky, R. Su, H. Nguyen, J. Moore, Y. Lou, S. Bhadra, L. Forte, M. Anastasio, and W. Yang, "Full-view 3D imaging system for functional and anatomical screening of the breast," SPIE BiOS **10494**, 104942Y (2018).
32. S.M. Schoustra, D. Piras, R. Huijink, T. Op't Root, L. Alink, W.M. Kobold, W. Steenbergen, and S. Manohar, "Twente Photoacoustic Mammoscope 2: system overview and three-dimensional vascular network images in healthy breasts," J. Biomed. Opt. **24**(12), 1–12 (2019).
33. M. Toi, Y. Asao, and Y. Matsumoto, *et al.*, "Visualization of tumor-related blood vessels in human breast by photoacoustic imaging system with a hemispherical detector array," Sci. Rep. **7**(1), 41970 (2017).
34. K. Nagae, Y. Asao, and Y. Sudo, *et al.*, "Real-time 3D photoacoustic visualization system with a wide field of view for imaging human limbs," F1000Res **7**, 1813 (2018).
35. N. Nyayapathi, R. Lim, H. Zhang, W. Zheng, Y. Wang, M. Tiao, K.W. Oh, X.C. Fan, E. Bonaccio, K. Takabe, and J. Xia, "Dual Scan Mammoscope (DSM)—a new portable photoacoustic breast imaging system with scanning in craniocaudal plane," IEEE Trans. Biomed. Eng. **67**(5), 1321–1327 (2020).
36. C. Lee, W. Choi, J. Kim, and C. Kim, "Three-dimensional clinical handheld photoacoustic/ultrasound scanner," Photoacoustics **18**, 100173 (2020).

Biomedical Optics EXPRESS

37. N. Holzwarth, M. Schellenberg, J. Grohl, K. Dreher, J.H. Nolke, A. Seitel, M.D. Tizabi, B.P. Muller-Stich, and L. Maier-Hein, "Tattoo tomography: Freehand 3D photoacoustic image reconstruction with an optical pattern," Int. J. Comput. Assist. Radiol. Surg. **16**(7), 1101–1110 (2021).
38. B. Park, C.H. Bang, C. Lee, J.H. Han, W. Choi, J. Kim, G.S. Park, J.W. Rhie, J.H. Lee, and C. Kim, "3D wide-field multispectral photoacoustic imaging of human melanomas in vivo: a pilot study," J. Eur. Acad. Dermatol. Venereol. **35**(3), 669–676 (2021).
39. B. Wang, C. Wang, F. Zhong, W. Pang, L. Guo, K. Peng, and J. Xiao, "3D acoustic resolution-based photoacoustic endoscopy with dynamic focusing," Quant. Imaging Med. Surg. **11**(2), 685–696 (2021).
40. N. Gandhi, S. Kim, P. Kazanzides, and M.A. Lediju Bell, "Accuracy of a novel photoacoustic-based approach to surgical guidance performed with and without a da Vinci robot," in Photons Plus Ultrasound: Imaging and Sensing (2017).
41. A. Benjamin, M. Chen, Q. Li, C.A. Carrascal, H. Xie, A. Samir, and B. Anthony, "Renal volume reconstruction using free-hand ultrasound scans," J. Acoust. Soc. Am. **145**(3), 1922 (2019).
42. P. Hausamann, C.B. Sinnott, M. Daumer, and P.R. MacNeilage, "Evaluation of the Intel RealSense T265 for tracking natural human head motion," Sci. Rep. **11**(1), 12486 (2021).
43. A. Benjamin, M. Chen, Q. Li, L. Chen, Y. Dong, C.A. Carrascal, H. Xie, A.E. Samir, and B.W. Anthony, "Renal volume estimation using freehand ultrasound scans: an ex vivo demonstration," Ultrasound Med. Biol. **46**(7), 1769–1782 (2020).
44. Datasheet: Intel RealSense Tracking Camera T265, 2019; Available from: https://www.intelrealsense.com/wp-content/uploads/2019/09/Intel_RealSense_Tracking_Camera_Datasheet_Rev004_release.pdf?_ga=2.42986661.1424111992.1669915299-653277717.1669915299.
45. F. Vulpi, R. Marani, A. Petitti, G. Reina, and A. Milella, "An RGB-D multi-view perspective for autonomous agricultural robots," Computers and Electronics in Agriculture **202**, 107419 (2022).
46. D. Jiang, H. Chen, R. Zheng, and F. Gao, "Hand-held free-scan 3D photoacoustic tomography with global positioning system," J. Appl. Phys. **132**(7), 074904 (2022).
47. M.O.A. Aqel, M.H. Marhaban, M.I. Saripan, and N.B. Ismail, "Review of visual odometry: types, approaches, challenges, and applications," SpringerPlus **5**(1), 1897 (2016).
48. C. Peng, Q. Cai, M. Chen, and X. Jiang, "Recent advances in tracking devices for biomedical ultrasound imaging applications," Micromachines **13**(11), 1855 (2022).
49. S.-Y. Sun, M. Gilbertson, and B.W. Anthony, *Probe Localization for Freehand 3D Ultrasound by Tracking Skin Features* (Springer International Publishing, 2014).
50. A. Alapetite, Z. Wang, J.P. Hansen, M. Zajączkowski, and M. Patalan, "Comparison of three off-the-shelf visual odometry systems," Robotics **9**(3), 56 (2020).
51. Y. Zhu, G. Xu, J. Yuan, J. Jo, G. Gandikota, H. Demirci, T. Agano, N. Sato, Y. Shigeta, and X. Wang, "Light emitting diodes based photoacoustic imaging and potential clinical applications," Sci. Rep. **8**(1), 9885 (2018).